WHAT IS CLAIMED:

1. A method for use in the analysis of gene or protein expression information comprising,

(a) accessing gene or protein expression data comprising expression levels of G genes or proteins in S samples, where the S samples may be classified into C classes representing cellular states;

(b) determining a measure of the variability of expression levels of each gene or protein in the data as a whole; and

(c) determining a measure of the variability of expression levels of each gene or protein within each class of sample.

2. The method of claim 1, further comprising:

(d) determining a between group measure of variability by determining the difference between the measure of variability determined in (c) from the measure of variability determined in (b).

3. The method of claim 1, further comprising generating a comparison of the measure of variability determined in (b) to the measures of variability determined in (c).

4. The method of claim 2, further comprising generating a comparison of the measure of variability determined in (c) to the measure of variability determined in (d).

5. The method of claim 1, wherein the comparison comprises determining the ratio of the measure of variability of (c) to the measure of variability of (b).

6. The method of claim 2, wherein the comparison comprises determining the ratio of the measure of variability of (c) to the measure of variability of (d).

7. The method of claim 3, wherein the comparison comprises calculating a Wilks' lambda score.

8. The method of claim 3, wherein the comparison comprises scaling the measure of variability of (b) by the measure of variability of (c).

9. The method of claim 1, wherein the measure of variability is selected from the group consisting of: variance and kurtosis.

10. The method of claim 1, wherein G is one.

11. The method of claim 1, wherein G is two or greater.

12. The method of claim 1, wherein C is two or greater, and wherein S is equal to or greater than C.

13. The method of claim 1, wherein the data is organized into a data matrix X_k for each class k, and wherein each data matrix is organized such that $X(i,j)$ is the expression of gene j in sample i.

14. The method of claim 2, wherein the measure of variability determined in (c) is represented by a matrix $\boldsymbol{B}$, and wherein the measure of variability determined in (d) is represented by a matrix $\boldsymbol{W}$.

15. The method of claim 14, wherein $\boldsymbol{W}$ is generated according to the formula

$$\boldsymbol{W} = \sum_{k=1}^{c} \left(X_k - 1\bar{x}_k\right)^T \left(X_k - 1\bar{x}_k\right)$$

wherein $\bar{\mathbf{x}}_k$ is the group mean ($1 \times g$) for class k.

16. The method of claim 14, wherein $\boldsymbol{B}$ is generated according to the formula

$$\boldsymbol{B} = \boldsymbol{T} - \boldsymbol{W} = \left(\mathbf{X} - 1\bar{\mathbf{x}}\right)^T \left(\mathbf{X} - 1\bar{\mathbf{x}}\right) - \mathbf{W}$$

wherein $\bar{\mathbf{x}}_k$ is the group mean ($1 \times g$) for class k, $\bar{\mathbf{x}}$ is the mean for all the data, and $\boldsymbol{T}$ is the total variance of all the data..

17. The method of claim 14, comprising generating a comparison of the matrix $\boldsymbol{B}$ and the matrix $\boldsymbol{W}$.

18. The method of claim 17, wherein the comparison is a matrix $\boldsymbol{W}^{-1}\boldsymbol{B}$.

19. The method of claim 17, further comprising maximizing the separation between the classes in a reduced dimensional space.

20. The method of claim 19, wherein maximizing the separation between the classes in a reduced dimensional space comprises generating an eigenvector matrix $\boldsymbol{L}$ of the matrix $\boldsymbol{W}^{1}\boldsymbol{B}$ and an eigenvalue matrix $\boldsymbol{\Lambda}$ of the matrix $\boldsymbol{W}^{1}\boldsymbol{B}$.

21. The method of claim 20, wherein a column of $\boldsymbol{L}$ defines a discriminant function of the reduced dimensional space, and wherein each entry in the column indicates the contribution of each gene to the discriminant function.

22. The method of claim 19, wherein the variance-covariance structure is similar in each class.

23. The method of claim 19, wherein maximizing the separation between the classes in a reduced dimensional space comprises generating a singular value decomposition of the matrix $\boldsymbol{W}^{1}\boldsymbol{B}$.

24. The method of claim 23, wherein generating a singular value decomposition of the matrix $\boldsymbol{W}^{1}\boldsymbol{B}$ is performed according to the formula:

$$\mathbf{W}^{-1}\mathbf{B} = \mathbf{U}\Lambda\mathbf{L}^{T}$$

wherein $\boldsymbol{U}$ is a left singular vector, $\boldsymbol{L}$ is a matrix of discriminant functions, and $\boldsymbol{\Lambda}$ is a matrix of singular values representing the discriminant loadings in the corresponding functions.

25. The method of claim 21, further comprising calculating a discriminator vector for each sample i, wherein the discriminator vector represents a position of the sample in the reduced dimensional space.

26. The method of claim 25, wherein calculating a discriminator vector comprises operating the formula:

$$y_j = \boldsymbol{i}\boldsymbol{L}_j = \sum_{z=1}^{g} i_z L_{ij}$$

wherein y_j is the discriminator score of the sample i (a sample of g genes) for each column j of matrix L , and wherein the discriminator vector is a combination of each y_j into a vector having a dimensionality that is equal to the number of dimensions in the reduced dimensional space.

27. The method of claim 3, further comprising generating discriminant loadings based on the comparison.

28. The method of claim 4, further comprising generating discriminant loadings based on the comparison.

29. The method of claim 27, further comprising generating a discriminator vector for each sample, wherein the discriminator vector describes a point in a space having one or more dimensions.

30. The method of claim 28, further comprising generating a discriminator vector for each sample based on the comparison, wherein the discriminator vector describes a point in a space having one or more dimensions.

31. The method of claim 27, further comprising determining the contribution of the expression levels of a gene or protein to the discriminant loadings, wherein a gene or protein that contributes significantly to a dimension is a gene or protein that is related to a cellular state of one or more sample.

32. The method of claim 28, further comprising determining the contribution of the expression levels of a gene or protein to the discriminant loadings, wherein a gene or protein that contributes significantly to a dimension is a gene or protein that is related to a cellular state or a change in cellular state of one or more sample.

33. The method of claim 31, wherein C is two or greater and the space has C-1 dimensions.

34. The method of claim 32, further comprising generating a rank order list of genes or proteins based on contribution to the dimensions of the space.

35. The method of claim 34, wherein the rank order list is generated by comparing the F score for each gene or protein.

36. A method for identifying a gene or protein, the expression of which is related to a cellular state or a change in cellular state comprising,

(a) accessing gene or protein expression data comprising expression levels of G genes or proteins in S samples, where the S samples may be classified into C classes representing cellular states;

(b) determining a measure of the variability of expression levels of each gene or protein in the data as a whole;

(c) determining a measure of the variability of expression levels of each gene or protein within each class of sample; and

(d) identifying a gene or protein which is related to a cellular state or a change in cellular state by identifying a gene or protein for which the measure of variability determined in (c) is less than the measure of variability determined in (b) with a 90% degree of confidence.

37. The method of claim 36, wherein C is two or greater.

38. The method of claim 36, wherein the measure of variability is selected from the group consisting of: variance and kurtosis.

39. The method of claim 36, wherein (d) comprises performing a Fisher Discriminant Analysis.

40. The method of claim 36, further comprising identifying a plurality of genes or proteins according to (d), and wherein the plurality of genes or proteins is a gene or protein group related to a cellular state or a change in a cellular state.

41. The method of claim 36, wherein the gene or protein group is analyzed by determining the contribution of each gene or protein of the group to the power of the group to discriminate between two or more classes of sample.

42. The method of claim 41, wherein determining the contribution of a gene or protein of interest comprises,

(i) generating a subgroup by omitting the gene or protein of interest from the group; and

(ii) testing the power of the subgroup to discriminate between two or more classes of sample.

43. The method of claim 41, wherein determining the contribution of a gene or protein comprises, performing a leave one out cross-validation.

44. A method for identifying a gene or protein expression pattern that is useful for discriminating between samples of two or more cellular states comprising,

(a) accessing gene or protein expression data comprising expression levels of G genes or proteins in S samples, where the S samples may be classified into C classes representing cellular states;

(b) determining a measure of the variability of expression levels of each gene or protein in the data as a whole;

(c) determining a measure of the variability of expression levels of each gene or protein within each class of sample;

(d) generating for each gene or protein a comparison of the measure of variability determined in (b) to the measures of variability determined in (c); and

(e) selecting from among the genes or proteins of (d) a set of genes or proteins and corresponding expression levels that discriminate between two or more classes of sample with a

misclassification rate less than 40%, wherein the set of genes or proteins and corresponding expression levels is a pattern that is useful for discriminating between samples of two or more cellular states.

45. The method of claim 44, wherein the pattern is useful for discriminating between cellular states selected from the group consisting of:

(a) hyperprliferative and non-hyperproliferative epithelial cells

(b) AML, B-ALL and T-ALL; and

(c) a bacterium producing higher levels of a metabolite and a bacterium producing lower levels of a metabolite.

46. A computer product for use in analyzing gene or protein expression data, the product disposed on a computer readable medium, and comprising instructions for causing a processor to:

(a) determine a measure of the variability of expression levels of a gene or protein in gene or protein expression data comprising expression levels of G genes or proteins in S samples, where the S samples may be classified into C classes representing cellular states;

(b) determining a measure of the variability of expression levels of the gene or protein within each class of sample in the data.

47. The computer product of claim 46, further comprising instructions for causing a processor to generate for the gene or protein a comparison of the measure of variability determined in (b) to the measures of variability determined in (c).

48. A system comprising a processor and instructions for causing a processor to:

(a) determine a measure of the variability of expression levels of each gene or protein in gene or protein expression data comprising expression levels of G genes or proteins in S samples, where the S samples may be classified into C classes representing cellular states;

(b) determining a measure of the variability of expression levels of each gene or protein within each class of sample in the data.

49. A method for use in modifying the production of a metabolite in a cell comprising:

(a) accessing data comprising a representation of the expression levels of G genes or proteins in S samples, wherein the S samples may be classified into C classes representing biological states, and wherein at least two of the biological states differ in the level of the metabolite that is produced; and

(b) identifying a discriminating gene or protein, the expression levels of which are discriminatory in defining a biological state of higher metabolite production from a biological state of lower metabolite production.

50. The method of claim 49, further comprising identifying a discriminating pattern of gene or protein expression levels.

51. The method of claim 50, further comprising selecting a cell having a desired level of metabolite production by identifying a cell having a pattern of expression levels that is mathematically similar to the discriminating pattern.

52. The method of claim 49, further comprising modulating the expression of the candidate gene or protein in a cell.

53. The method of claim 49, wherein identifying a gene or protein comprises:

(i) determining a measure of the variability of expression levels of each gene or protein in the data as a whole; and

(ii) determining a measure of the variability of expression levels of each gene within each class of sample.

54. A method for use in modifying the production of a polyhydroxyalkanoate in a cell comprising altering the genetic makeup of the cell so as to cause the cell to have a modified expression of a gene represented by an index number selected from the group consisting of: sll0008, sll0010, sll0039, sll0322, sll0361, sll0373, sll0374, sll0379, sll0385, sll0396, sll0459, sll0469, sll0477, sll0486, sll0550, sll0558, sll0703, sll0873, sll1317, sll1376, sll1473, sll1504, sll1514, sll1611, sll1623, sll1630, sll1632, sll1702, sll1820 and slr1822, or an orthologue of any of the preceding.

55. A method of claim 54, wherein altering the genetic makeup of the cell comprises introducing a recombinant nucleic acid into the cell.

56. A method of claim 54, wherein the cell is a cyanobacterial cell.

57. A method of claim 54, wherein the cell is a bacterium selected from the group consisting of: *Synechocystis sp.*, *Synechococcus sp.*, *Ralstonia eutropha*, *Alcaligenes latus*, *Azotobacter vinelandii*, *Anacystis nidulans* and recombinant *Escherichia coli*.

58. A method of claim 54, wherein the polyhydoxyalkanoate is selected from the group consisting of: polyhydroxyproprionate, polyhydroxybutyrate, polyhydroxyvalerate, polyhydroxycaproate, polyhydroxyheptanoate, polyhydroxyoctanoate, polyhydroxynonanoate, polyhydroxydecanoate, polyhydroxyundecanoate, polyhydroxydodecanoate and a mixed polymer of one or more of the forgoing polymers.

59. A bacterium comprising a recombinant nucleic acid construct comprising a coding sequence of a gene represented by an index number selected from the group consisting of: sll0008, sll0010, sll0039, sll0322, sll0361, sll0373, sll0374, sll0379, sll0385, sll0396, sll0459, sll0469,

sll0477, sll0486, sll0550, sll0558, sll0703, sll0873, sll1317, sll1376, sll1473, sll1504, sll1514, sll1611, sll1623, sll1630, sll1632, sll1702, sll1820 and slr1822, or an orthologue of any of the preceding.

60. The bacterium of claim 59, wherein the bacterium is selected from the group consisting of: *Synechocystis sp.*, *Synechococcus sp.*, *Ralstonia eutropha*, *Alcaligenes latus*, *Azotobacter vinelandii*, *Anacystis nidulans* and *Escherichia coli*.

61. A method of producing a polyhydroxyalkanoate comprising:

(a) growing a culture of cells of claim 59 under conditions suitable for the production of a polyhydroxyalkanoate ; and

(b) obtaining a polyhydroxyalkanoate from the culture.

62. A method of claim 61, further comprising refining the polyhydroxyalkanoate to obtain a purer form of polyhydroxyalkanoate.

63. A method for determining whether a sample contains a hyperproliferative cell comprising:

a) determining a level of gene expression of at least one gene in a sample, wherein the at least one gene is selected from the group consisting of Neuromedin U; Aldehyde dehydrogenase 9 (Human gamma-aminobutyraldehyde dehydrogenase E3 isozyme); Fibroblast growth factor 8; Human epidermal growth factor receptor (HER3); Translocase of outer mitochondrial membrane 34; KIAA0089; Monoamine oxidase B; Zinc finger protein 273; clone 1D2; Aldehyde dehydrogenase 10 (fatty aldehyde dehydrogenase); Carboxylesterase 2 (intestine, liver); Gro2 oncogene; Diazepam binding inhibitor; Cadherin 17; TAL1 (SCL) interrupting locus; Crystallin alpha B; 5T4 oncofetal trophoblast glycoprotein; Deoxyribonuclease I-like 3; Heat-shock protein 90-kDa; Smg GDS-associated protein; Cytochrome c oxidase subunit Vb (coxVb); Wilm Tumor-

Related Protein; TYRO3 protein tyrosine kinase; FAT tumor suppressor; Creatine kinase, mitochondrial 1; Transcription factor 20; MHC class I polypeptide related sequence A; KIAA0018 gene product 1; Lectin galactoside-binding, soluble, 7 (galectin 7); Tenascin-R (restrictin, janusin); CD1A antigen, a polypeptide; Beta-Hexosaminidase, Alpha Polypeptide, Abnormal Splice Mutation; clone 1A7; KIAA0172 gene; Myxovirus (influenza) resistance 2, homolog of murine; Lysophospholipase like; Interleukin-8 receptor type B, splice variant IL8RB9; keratin 4; and Runt-related transcription factor, and wherein the level of gene expression of the at least one gene allows classification of an oral keratinocyte as hyperproliferative or non-hyperproliferative with a misclassification rate of 40% or lower;

b) comparing the level of gene expression of said at least one gene to a first control level of gene expression of said at least one gene as measured in a hyperproliferative cell; and

c) comparing the level of gene expression of the at least one gene to a second control level of gene expression of said at least one gene as measured in a non-hyperproliferative cell;

wherein a sample contains a hyperproliferative cell if the level of gene expression of the at least one gene is more mathematically similar to the first control level of gene expression than to the second control level of gene expression.

64. The method of claim 63, wherein the expression levels of at least 2 genes are determined and compared in steps (a), (b), and (c).

65. The method of claim 63, wherein the misclassification rate is 15% or lower.

66. The method of claim 63, wherein the hyperproliferative cell is a cancer cell.

67. The method of claim 63, wherein the hyperproliferative cell is an oral cancer cell.

68. A method for determining whether a sample contains a hyperproliferative, cell comprising:

a) determining a level of gene expression of at least two genes in a sample, wherein said at least two genes are selected from the group consisting of Neuromedin U; Aldehyde dehydrogenase 9 (Human gamma-aminobutyraldehyde dehydrogenase E3 isozyme); Fibroblast growth factor 8; Human epidermal growth factor receptor (HER3); Translocase of outer mitochondrial membrane 34; KIAA0089; Monoamine oxidase B; Urokinase plasminogen activator; Zinc finger protein 273; clone 1D2; Aldehyde dehydrogenase 10 (fatty aldehyde dehydrogenase); Carboxylesterase 2 (intestine, liver); Gro2 oncogene; Diazepam binding inhibitor; Cadherin 17; TAL1 (SCL) interrupting locus; Crystallin alpha B; 5T4 oncofetal trophoblast glycoprotein; Deoxyribonuclease I-like 3; Heat-shock protein 90-kDa; Smg GDS-associated protein; Cytochrome c oxidase subunit Vb (coxVb); Wilm Tumor-Related Protein; TYRO3 protein tyrosine kinase; FAT tumor suppressor; Creatine kinase, mitochondrial 1; Ferritin, light polypeptide; Transcription factor 20; MHC class I polypeptide related sequence A; KIAA0018 gene product 1; Lectin galactoside-binding, soluble, 7 (galectin 7); Tenascin-R (restrictin, janusin); CD1A antigen, a polypeptide; Cytochrome P4502C9 subfamily IIC (mephytoin4-hydroxylase), polypeptide 9; Phospholipase A2, group VII; Beta-Hexosaminidase, Alpha Polypeptide, Abnormal Splice Mutation; clone 1A7; KIAA0172 gene; Interleukin 8 receptor, beta; Myxovirus (influenza) resistance 2, homolog of murine; Lysophospholipase like; Interleukin-8 receptor type B, splice variant IL8RB9; keratin 4; Runt-related transcription factor; and Cathepsin L; and wherein the level of gene expression of said at least two genes allows classification of an oral keratinocyte as

hyperproliferative or non-hyperproliferative with a misclassification rate of 40% (30%, 20%, 15%, 10%) or lower;

b) comparing the level of gene expression of said at least two genes to a first control level of gene expression of said at least two genes as measured in a hyperproliferative cell; and

c) comparing the level of gene expression of the at least two genes to a second control level of gene expression of said at least two genes as measured in a non-hyperproliferative cell;

wherein a sample contains a hyperproliferative cell if the level of gene expression of the at least one gene is more mathematically similar to the first control level of gene expression than to the second control level of gene expression.

69. A method for classifying a leukemia sample comprising:

a) determining a level of gene expression of at least one gene in a sample, wherein said at least one gene is selected from the group consisting of U05259, M89957, M84371, D88270, X58529, M28170, M31523, M11722, JO3473, X03934, U23852, X00437, M23323, X59871, X76223, D00749, L05148, U14603, M37271, M26692, M12886, J05243, X69398, U67171, X04145, L10373, U16954, J04132, M28826, HG4128, X87241, U50743, M13792, L47738, X95735, X17042, M23197, M84526, L09209, U46499, M27891, M16038, M63138, M55150, M22960, M62762, X61587, and U50136, and wherein the level of gene expression of said at least one gene allows classification of a leukemia as AML, B-ALL or T-ALL with a misclassification rate of 40% or lower;

b) comparing the level of gene expression of said at least one gene to a first control level of gene expression of said at least one gene as measured in an AML cell;

c) comparing the level of gene expression of said at least one gene to a second control level of gene expression of said at least one gene as measured in a B-ALL cell; and

d) comparing the level of gene expression of said at least one gene to a third level of gene expression of said at least one gene as measured in a T-ALL cell;

wherein the leukemia is classified as AML, B-ALL or T-ALL depending on whether the level of gene expression of the at least one gene is more mathematically similar to the first control level of gene expression; the second control level of gene expression; or the third control level of gene expression.

70. The method of claim 69, wherein the expression levels of at least 2 genes are determined and compared in steps (a), (b), (c) and (d).

71. The method of claim 69, wherein the misclassification rate is 15% or lower.

72. A method for classifying a leukemia sample comprising:

a) determining a level of gene expression of at least one gene in a sample, wherein said at least one gene is selected from the group consisting of M89957, M84371, D88270, X58529, M28170, M11722, JO3473, X03934, U23852, X00437, M23323, X59871, X76223, D00749, L05148, U14603, M37271, M26692, M12886, J05243, X69398, U67171, X04145, L10373, U16954, J04132, M28826, HG4128, X87241, U50743, L09209, U46499, M22960, and X61587, and wherein the level of gene expression of said at least one gene allows classification of a leukemia as AML, or ALL with a misclassification rate of 40% or lower;

b) comparing the level of gene expression of said at least one gene to a first control level of gene expression of said at least one gene as measured in an AML cell;

c) comparing the level of gene expression of said at least one gene to a second control level of gene expression of said at least one gene as measured in an ALL cell; and wherein the leukemia is classified as AML or ALL depending on whether the level of gene expression of the at least one gene is more mathematically similar to the first control level of gene expression or the second control level of gene expression.

73. A method for identifying a candidate therapeutic agent for the treatment of a hyperproliferative disorder comprising:

(a) contacting a hyperproliferative cell with a test therapeutic agent;

(b) determining a level of gene expression of a gene in the cell, wherein said gene is selected from the group consisting of Neuromedin U; Aldehyde dehydrogenase 9 (Human gamma-aminobutyraldehyde dehydrogenase E3 isozyme); Fibroblast growth factor 8; Human epidermal growth factor receptor (HER3); Translocase of outer mitochondrial membrane 34; KIAA0089; Monoamine oxidase B; Urokinase plasminogen activator; Zinc finger protein 273; clone 1D2; Aldehyde dehydrogenase 10 (fatty aldehyde dehydrogenase); Carboxylesterase 2 (intestine, liver); Gro2 oncogene; Diazepam binding inhibitor; Cadherin 17; TAL1 (SCL) interrupting locus; Crystallin alpha B; 5T4 oncofetal trophoblast glycoprotein; Deoxyribonuclease I-like 3; Heat-shock protein 90-kDa; Smg GDS-associated protein; Cytochrome c oxidase subunit Vb (coxVb); Wilm Tumor-Related Protein; TYRO3 protein tyrosine kinase; FAT tumor suppressor; Creatine kinase, mitochondrial 1; Ferritin, light polypeptide; Transcription factor 20; MHC class I polypeptide related sequence A; KIAA0018 gene product 1; Lectin galactoside-binding, soluble, 7 (galectin 7); Tenascin-R (restrictin, janusin); CD1A antigen, a polypeptide; Cytochrome P4502C9 subfamily IIC (mephytoin4-

hydroxylase), polypeptide 9; Phospholipase A2, group VII; Beta-Hexosaminidase, Alpha Polypeptide, Abnormal Splice Mutation; clone 1A7; KIAA0172 gene; Interleukin 8 receptor, beta; Myxovirus (influenza) resistance 2, homolog of murine; Lysophospholipase like; Interleukin-8 receptor type B, splice variant IL8RB9; keratin 4; Runt-related transcription factor; and Cathepsin L; and wherein the level of gene expression of said gene allows classification of an oral keratinocyte as hyperproliferative or non-hyperproliferative with a misclassification rate of 40% or lower; and

(c) determining whether the expression level of said gene is more mathematically similar to that of a proliferative cell or a non-hyperproliferative cell, wherein a test therapeutic agent that causes the expression level of the gene in the hyperproliferative cell to more closely resemble the expression level of the gene in a non-hyperproliferative cell is a candidate therapeutic agent.